

ALGORITHMIC THEORY OF FREE SOLVABLE GROUPS: RANDOMIZED COMPUTATIONS

ALEXANDER USHAKOV

ABSTRACT. We design new deterministic and randomized algorithms for computational problems in free solvable groups. In particular, we prove that the word problem and the power problem can be solved in quasi-linear time and the conjugacy problem can be solved in quasi-quartic time by Monte Carlo type algorithms.

Keywords. Solvable groups, metabelian groups, word problem, cyclic subgroup membership, power problem, conjugacy problem, randomized algorithms.
2010 Mathematics Subject Classification. 03D15, 20F65, 20F10.

1. INTRODUCTION

The study of algorithmic problems in free solvable groups can be traced to the work [11] of Magnus, who in 1939 introduced an embedding (now called the *Magnus embedding*) of an arbitrary group of the type $F/[N, N]$ into a matrix group of a particular type with coefficients in the group ring of F/N (see Section 1.5 below). Since the word problem in free abelian groups is decidable in polynomial time, by induction, this embedding gives a polynomial time decision algorithm for a fixed free solvable group $S_{r,d}$. However the degree of the polynomial here grows together with d . An algorithm polynomial in both: the length of a given word and the class d of the free solvable group was found later in [15]. It was proved that the word problem has time complexity $O(r|w| \log_2 |w|)$ in the free metabelian group $S_{r,2}$, and $O(rd|w|^3)$ in a free solvable group $S_{r,d}$ for $d \geq 3$.

The general approach to the conjugacy problem in wreath products was suggested by Matthews in [14] who also described the solution to the conjugacy problem in free metabelian groups. The first solution to the conjugacy problem in free solvable groups was given by Remeslennikov and Sokolov in [20] who proved that the conjugacy in $S_{r,d}$ can be reduced to the conjugacy in a wreath product of $S_{r,d-1}$ and a free abelian group. Later Vassileva showed in [24] that the power problem in free solvable groups can be solved in $O(rd(|u| + |v|)^6)$ time and used that result to show that the Matthews-Remeslennikov-Sokolov approach can be transformed into a polynomial time $O(rd(|u| + |v|)^8)$ algorithm. In this paper we improve the results of [15] and [24], namely we prove that:

Theorem 2.6. There exists a quasi-quadratic time $\tilde{O}(|w|^2)$ deterministic algorithm solving the word problem in $S_{r,d}$.

Theorem 5.1. There exists a quasi-quadratic time $\tilde{O}((|u| + |v|)^2)$ deterministic algorithm solving the power problem in $S_{r,d}$.

The author was partially supported by NSF grant DMS-0914773 and by NSA Mathematical Sciences Program grant number H98230-14-1-0128.

The author would like to thank Andrey Nikolaev for his helpful and insightful comments.

Theorem 6.5. There exists a quasi-quintic time $\tilde{O}((|u| + |v|)^5)$ deterministic algorithm solving the conjugacy problem in $S_{r,d}$.

We can improve these results further if we grant our machine an access to a random number generator. The price of that improvement is an occasional incorrectness of the result. Fortunately, we can control the probability of an error: for any fixed polynomial p we can adjust some internal parameter in the algorithm to guarantee that the probability of an error converges to 0 as fast as $O(1/p(n))$.

Theorem 4.5. There exists a quasi-linear time $\tilde{O}(|w|)$ false-biased randomized algorithm solving the word problem in $S_{r,d}$. \square

Theorem 5.2. There exists a quasi-linear time $\tilde{O}(|u| + |v|)$ unbiased randomized algorithm solving the power problem in $S_{r,d}$. \square

Theorem 6.6. There exists a quasi-quartic time $\tilde{O}((|u| + |v|)^4)$ unbiased randomized algorithm solving the conjugacy problem in $S_{r,d}$. \square

Also, we want to mention Theorem 6.4 which gives a geometric approach to the conjugacy problem in free solvable groups.

Theorem 6.4. Words $x, y \in F(X_r)$ represent conjugate elements in $S_{r,d}$ if and only if there exists $z \in F(X_r)$ such that zxz^{-1} and y define the same flows in the Schreier graph of $\langle y \rangle$ in $S_{r,d-1}$. \square

1.1. Randomized algorithms. A randomized algorithm is an algorithm which uses randomness as a part of its logic. Typically it uses uniformly random bits as an auxiliary input to guide its behavior in the hope of achieving good performance in the average case over all possible choices of random bits.

Historically, the first randomized algorithm was a method developed by M. Rabin in [18] for the closest pair problem in computational geometry. The study of randomized algorithms was spurred by the 1977 discovery of a randomized primality test by R. Solovay and V. Strassen in [22]. Soon afterwards M. Rabin in [19] demonstrated that the Miller's primality test can be turned into a very efficient $\tilde{O}(\log^2(n))$ randomized algorithm. At that time, no practical deterministic algorithm for primality was known. Even though a deterministic polynomial-time $\tilde{O}(\log^6(n))$ primality test has since been found (see AKS primality test, [1]), it has not replaced the older probabilistic tests in cryptographic software nor is it expected to do so for the foreseeable future. See [16] for more on randomized algorithms. There are two main types of randomized algorithms: Las Vegas and Monte Carlo algorithms.

A *Monte Carlo algorithm* is a randomized algorithm whose running time is deterministic, but whose output may be incorrect with a certain (typically small) probability. For decision problems, these algorithms are generally classified as either false-biased or true-biased. A false-biased Monte Carlo algorithm is always correct when it returns false; a true-biased behaves likewise. While this describes algorithms with one-sided errors, others might have no bias; these are said to have two-sided errors. The answer they provide (either true or false) will be incorrect, or correct, with some bounded probability. The Solovay-Strassen primality test always answers true for prime number inputs; for composite inputs, it answers false with probability at least $1/2$ and true with probability at most $1/2$. Thus, false answers from the algorithm are certain to be correct, whereas the true answers remain uncertain; this is said to be a $(1/2)$ -correct false-biased algorithm.

A *Las Vegas algorithm* is a randomized algorithm that always gives correct results; that is, it always produces the correct result or it informs about the failure. Las Vegas algorithms were introduced by László Babai in 1979, in the context of the graph isomorphism problem, as a stronger version of Monte Carlo algorithms, see [2].

1.2. Algorithmic problems in groups. Let $F = F_r = F(X)$ be a free group with a basis $X = X_r = \{x_1, \dots, x_r\}$. By $|w|$ we denote the *length* of $w \in F$. By ε we denote the empty word. When $|uv| = |u| + |v|$, then we write $u \circ v$ for uv . Let $R \subseteq F$. A pair (X, R) defines a *presentation* of a group $G = F/N$ (also denoted by $\langle X \mid R \rangle$), where $N = \mathbf{ncl}(R)$ is the *normal closure* of R in F . If R is finite [recursively enumerable], then the presentation is called finite [recursively enumerable]. For a recursively presented group G one can study the following algorithmic questions.

The word problem (WP) in $G = \langle X \mid R \rangle$: Given $w \in F(X)$ decide if $w = 1$ in G , or not.

The power problem (PP) in $G = \langle X \mid R \rangle$: Given $u, v \in F(X)$ compute $k \in \mathbb{Z}$ such that $u = v^k$ in $\langle X \mid R \rangle$. If such k does not exist, then return *Fail*.

The conjugacy problem (CP) in $G = \langle X \mid R \rangle$: Given $u, v \in F(X)$ decide if there exists $c \in F(X)$ satisfying $u = c^{-1}vc$, or not.

It is easy to see that decidability/complexity of problems above does not depend on the generating set X . See [12, 10] for more on algorithmic problems in groups.

1.3. X -digraphs. An X -labeled directed graph Γ (or an X -digraph) is a pair of sets (V, E) where the set V is called the *vertex set* and the set $E \subseteq V \times V \times X$ is called the *edge set*. An element $e = (v_1, v_2, x) \in E$ designates an edge with the *origin* v_1 (also denoted by $\alpha(e)$), the *terminus* v_2 (also denoted by $\omega(e)$), labeled by x (also denoted by $\mu(e)$). We often use notation $v_1 \xrightarrow{x} v_2$ to denote the edge (v_1, v_2, x) . A *path* in Γ is a sequence of edges $p = e_1, \dots, e_k$ satisfying $\omega(e_i) = \alpha(e_{i+1})$ for every $i = 1, \dots, k-1$. The *origin* $\alpha(p)$ of p is the vertex $\alpha(e_1)$, the *terminus* $\omega(p)$ is the vertex $\omega(e_k)$, and the *label* $\mu(p)$ of p is the word $\mu(e_1)\dots\mu(e_k)$. We say that an X -digraph Γ is:

- *rooted* if it has a special vertex, called the root;
- *folded* (or deterministic) if for every $v \in V$ and $x \in X$ there exists at most one edge with the origin v labeled with x ;
- *complete* if for every $v_1 \in V$ and $x \in X$ there exists an edge $v_1 \xrightarrow{x} v_2$;
- *inverse* if with every edge $e = g_1 \xrightarrow{x} g_2$ Γ also contains the *inverse edge* $g_2 \xrightarrow{x^{-1}} g_1$, denoted by e^{-1} .

All X -digraphs in this paper are connected. A *morphism* of two rooted X -digraphs is a graph morphism which maps the root to the root and preserves labels. For more information on X -digraphs we refer to [23, 8].

Example 1.1. The Cayley graph of the group F/N , denoted by $\mathbf{Cay}(F/N)$, is an X -digraph (V, E) , where $V = F/N$ and

$$E = \{g \xrightarrow{x} gx \mid g \in F/N, x \in X^\pm\}.$$

It is an inverse folded complete graph. We always assume that the trivial element is the root of $\mathbf{Cay}(F/N)$.

Another important example of an X -digraph is the Schreier graph $\mathbf{Sch}_G(H)$ of a subgroup H of a group $G = F(X)/N$ defined as (V, E) :

$$V = \{Hg \mid g \in G\} \text{ and } E = \{Hg \xrightarrow{x} Hgx \mid g \in G, x \in X^\pm\}.$$

The coset H is the root of $\mathbf{Sch}_G(H)$. \square

Let Γ be an inverse X -digraph. Clearly, $(e^{-1})^{-1} = e$. Hence, the set of all edges can be split into a disjoint union $E = E^+ \sqcup E^-$ satisfying $(E^+)^{-1} = E^-$ and $(E^-)^{-1} = E^+$. The set E^+ is called a set of *positive edges* and the set E^- is called a set of *negative edges*.

The rank $r(\Gamma)$ of an inverse X -digraph Γ is defined as $|E^+| - |T|$, where T is any spanning subtree of Γ . The fundamental group $\pi_1(\Gamma)$ is the group of labels of all cycles at the root; it is naturally a subgroup of $F(X)$ of the rank $r(\Gamma)$ (see [8]).

1.4. Flows on X -digraphs. Let Γ be a deterministic inverse X -digraph with the root v . A flow on Γ is a function $f : E^+(\Gamma) \rightarrow \mathbb{Z}$ satisfying the following equality $\sigma(v) = 0$, where σ is:

$$\sigma(v) = \sum_{\alpha(e)=v} f(e) - \sum_{\omega(e)=v} f(e)$$

for all vertices $v \in V(\Gamma)$ except maybe two vertices s and t for which:

$$\sigma(s) = -\sigma(t) = 1.$$

The vertex s is called the **source** and the vertex t is called the **sink** of the flow f . If s and t are not defined, then f is called a **circulation**. In this paper the source is always the root v of Γ and, hence, if $\sigma(s) = 0$ then the sink is v as well.

Flows on deterministic connected inverse rooted X -digraphs can be defined by words in $F(X)$ and only by them as follows. For every word $w \in F(X)$ there exists at most one path p_w in Γ with the origin v labeled with w , called the *trace* of w in Γ . If p_w exists, then we can define the flow π_w of w on Γ which for every $e \in E^+$ counts the number of times the edge e is traversed minus the number of times the edge e^{-1} is traversed by p_w . It is also true that for every flow f on Γ there exists $w \in F(X)$ such that $f \equiv \pi_w$, see [15, Lemma 2.5].

1.5. Free solvable groups: tools and techniques. For a free group $F = F^{(0)} = F(X)$ of rank r denote by $F^{(1)} = [F^{(0)}, F^{(0)}]$ the *derived* subgroup of F , and by $F^{(d)} = [F^{(d-1)}, F^{(d-1)}]$ the d -th *derived subgroup* of F , $d \geq 2$. A free solvable group of rank r and class d is defined as follows:

- $S_{r,0} = F/F^{(0)}$ is a *trivial group* of rank r ,
- $S_{r,1} = F/F^{(1)}$ is a *free abelian group* of rank r ,
- $S_{r,2} = F/F^{(2)}$ is a *free metabelian group* of rank r , and
- in general, $S_{r,d} = F/F^{(d)}$ is a *free solvable group* of rank r and class d .

In the sequel we usually identify the set X with its canonical images in $S_{r,d}$. Note that any two consecutive groups in the list above are related to each other: $S_{r,i} = F/N$ and $S_{r,i+1} = F/[N, N]$, where $N = F^{(i)}$. Hence, naturally, every general technique for free solvable groups studies relations between groups of the type $F/[N, N]$ and F/N establishing an inductive step.

One of the most powerful approaches to study free solvable groups is via the Magnus embedding. Let $\mathbb{Z}F/N$ be the group ring of F/N with integer coefficients. By $\gamma : F \rightarrow F/N$ we denote the canonical factorization epimorphism, as well its

linear extension to $\gamma : \mathbb{Z}F \rightarrow \mathbb{Z}F/N$. Let T be a free (left) $\mathbb{Z}F/N$ -module of rank r with a basis $\{t_1, \dots, t_r\}$. Then the set of matrices:

$$M(F/N) = \left\{ \begin{pmatrix} g & t \\ 0 & 1 \end{pmatrix} \middle| g \in F/N, t \in T \right\}$$

forms a group with respect to the matrix multiplication. It is easy to see that the group $M(F/N)$ is isomorphic to the restricted wreath product $\mathbb{Z}^r \mathbf{wr} F/N$.

Theorem (Magnus embedding, [11]). *The homomorphism $\phi : F \rightarrow M(F/N)$ defined by*

$$x_i \mapsto \begin{pmatrix} x_i^\gamma & t_i \\ 0 & 1 \end{pmatrix}, \quad i = 1, \dots, r,$$

satisfies $\ker \phi = N'$. Therefore, ϕ induces a monomorphism $\phi : F/[N, N] \hookrightarrow M(F/N)$.

The Magnus embedding gives a solution to the word problem for free solvable groups. Using induction on the solvability class d gives a polynomial estimate $O(r^{d-1}|w|^{2d-1})$ on the complexity of the word problem, see [15, Section 2.2]

Another important technique for studying free solvable groups was introduced and studied by R. Fox in a sequence of papers [5, 6, 7, 3] who invented *free differential calculus*. Recall that a *free partial derivative* $\frac{\partial w}{\partial x_i}$ of the element w of the group $F/[N, N]$ by x_i is an element of the group ring $\mathbb{Z}F/N$ given by the formula:

$$(1) \quad \frac{\partial w}{\partial x_i} = \sum_{1 \leq j \leq n, i_j=i, \varepsilon_j=1} x_{i_1}^{\varepsilon_1} \dots x_{i_{j-1}}^{\varepsilon_{j-1}} - \sum_{1 \leq j \leq n, i_j=i, \varepsilon_j=-1} x_{i_1}^{\varepsilon_1} \dots x_{i_j}^{\varepsilon_j}.$$

The following result is one of the principle technical tools in this area, it follows easily from the Magnus embedding theorem, but in the current form it is due to Fox.

Theorem (Fox). *Let N be a normal subgroup of F and $\gamma : \mathbb{Z}F \rightarrow \mathbb{Z}F/N$ the canonical epimorphism. Then for every $u \in F$ the following equivalence holds:*

$$\forall i \quad (\partial u / \partial x_i)^\gamma = 0 \iff u \in [N, N].$$

Another approach to study elements of groups $F/[N, N]$ comes from geometric flows on $\mathbf{Cay}(F/N)$. As we discussed in Section 1.3 a word $w \in F(X)$ determines a unique path p_w in $\mathbf{Cay}(F/N)$ labeled by w which starts at the root (the vertex corresponding to the identity of G). The path p_w further defines a geometric flow π_w on $\mathbf{Cay}(F/N)$. Figure 1 gives an example of a flow for a particular word w in $\mathbf{Cay}(F_2/[F_2, F_2])$. Nonzero values of π_w are shown on the edges and zero values are omitted.

Theorem ([4, 25, 15]). *Let N be a normal subgroup of F and $u, v \in F$. Then $u = v$ in $F/[N, N]$ if and only if $\pi_u = \pi_v$ in $\mathbf{Cay}(F/N)$. \square*

1.6. Computational model and data representation. All computations are assumed to be performed on a random access machine. (Quasi-)Linear time is very sensitive to the way one represents the data, so here we describe precisely how the inputs are given to us. We use base 2 positional number system in which presentations of integers are converted into integers via the rule:

$$(a_{k-1} \dots a_3 a_2 a_1 a_0)_2 = a_{k-1} 2^{k-1} + \dots + a_2 2^2 + a_1 2 + a_0,$$

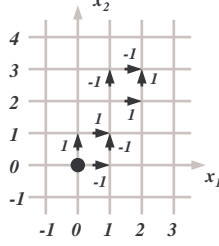


FIGURE 1. The values of π_w for $w = x_2 x_1 x_2 x_1 x_2 x_1^{-1} x_2^{-3} x_1^{-1}$ on (x_1, x_2) -grid. In this case $\partial w / \partial x_1 = -1 + x_2 - x_1 x_2^3 + x_1 x_2^2$ and $\partial w / \partial x_2 = 1 - x_1 + x_2^2 x_2^2 - x_1 x_2^2$.

where we assume that $a_{k-1} = 1$. The number k is called the *bit-length* of the presentation.

- Adding two numbers of bit-length at most n has $O(n)$ time complexity. The result is a number of bit-length at most $n + 1$.
- Computational complexity of multiplying two n -bit numbers is $O(n \log n \log \log n)$ (see [21]). The result is a $2n$ -bit number.

Let G be a group generated by a finite set $X_r = \{x_1, \dots, x_r\}$. We formally encode the word problem for G as a subset of $\{0, 1\}^*$ as follows. We first encode elements of the set $X_r^\pm = \{x_1^\pm, \dots, x_r^\pm\}$ by unique bit-strings of length $\lceil \log_2 r \rceil + 1$. The code for a word $w = w(X_r^\pm)$ is a concatenation of codes for letters and, formally:

$$\mathbf{WP}(S_{r,d}) = \{\text{code}(w) \mid w \in F(X_r), w = 1 \text{ in } S_{r,d}\}.$$

Thus, the bit-length of the representation for a word $w \in F(X_r)$ is:

$$|\text{code}(w)| = |w|(\lceil \log_2 r \rceil + 1).$$

We encode the power and conjugacy problems in a similar fashion. For both of these problems instances are pairs of words and the encoding can be done by introducing a new letter “,” into the alphabet X_r^\pm .

Note that any permutation of X_r induces an automorphism of a free solvable group and taking an automorphic image of a word preserves the property of being trivial. Furthermore, for any word w we can find in linear time in $|\text{code}(w)|$ an appropriate automorphic image satisfying $r \leq |w|$. Therefore, we always assume that $r \leq |w|$.

1.7. Quasi-linear time complexity. An algorithm is said to run in *quasi-linear time* if its time complexity function is $O(n \log^k n)$ for some constant $k \in \mathbb{N}$. We use notation $\tilde{O}(n)$ to denote quasi-linear time complexity. Quasi-linear time algorithms are also $o(n^{1+\varepsilon})$ for every $\varepsilon > 0$, and thus run faster than any polynomial in n with exponent strictly greater than 1. See [17] for more on quasi-linear time complexity theory. Similarly, one can define quasi-quadratic $\tilde{O}(n^2)$, quasi-cubic $\tilde{O}(n^3)$ time complexity as $O(n^2 \log^k n)$, $O(n^3 \log^k n)$, etc.

2. THE WORD PROBLEM: DETERMINISTIC SOLUTION

In this section we present a fast deterministic solution for the word problem in free solvable groups.

2.1. Support graphs. Let Γ be a rooted folded inverse X -digraph and m the length of a shortest cycle in Γ . Suppose that a reduced nontrivial word w can be traced in Γ . The set of edges traversed by w in Γ forms a connected X -digraph called the **support graph** of w in Γ .

Lemma 2.1. *Let Γ be a rooted folded inverse X -digraph and m the length of a shortest cycle in Γ . Suppose that a reduced nontrivial word w can be traced in Γ and $\pi_w = 0$. Then $|w| \geq 3m$.*

Proof. It follows from our assumption $\pi_w = 0$ that the path p_w is a cycle in Γ . Let Δ be the support graph of w in Γ . The rank of Δ can not be 0 (w is not reduced in this case) and can not be 1 (either w is not reduced or $\pi_w \neq 0$). Therefore, the rank of Δ is at least 2. Each edge of Δ is traversed by w at least twice. Hence, it is sufficient to prove that $2|E(\Delta)| \geq 3m$. Let Δ' be a minimal subgraph of Δ of rank exactly 2. There are exactly two distinct configurations possible for Δ' , shown in Figure 2.

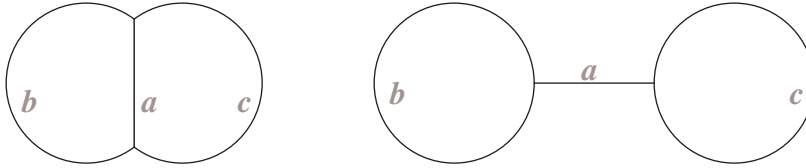


FIGURE 2. Two configurations for support graphs in Lemma 2.1.

Let a, b, c be the lengths of arcs as shown in the figure. Since, the length of a shortest cycle in Γ is m , we get the following bounds for our cases:

$$\begin{cases} a + b \geq m, \\ a + c \geq m, \\ b + c \geq m, \end{cases} \quad \begin{cases} b \geq m, \\ c \geq m. \end{cases}$$

In both cases we have $2(a + b + c) \geq 3m$ which proves that $2|E(\Delta)| \geq 3m$. Thus, $|w| \geq 3m$. \square

Proposition 2.2. *Let $w \in F(X_r) \setminus \{\varepsilon\}$. If $w = 1$ in $S_{r,d}$, then $|w| \geq 3^d$.*

Proof. Induction on d . The length of a shortest nontrivial relator in $S_{r,0}$ is 1. Assume that the statement holds for all values of solvability class less than d . Hence, the length of a shortest cycle in $\mathbf{Cay}(S_{r,d-1})$ is not smaller than 3^{d-1} . Choose a shortest nontrivial relator w in $S_{r,d}$. By Theorem 1.5 w defines the trivial flow in $\mathbf{Cay}(S_{r,d-1})$. By Lemma 2.1 $|w| \geq 3^d$. \square

2.2. Distinguishers. First, we fix some notation. For $a, b \in \mathbb{Z}$ ($a \leq b$) define a set:

$$[a, b] = \{a, a+1, \dots, b-1, b\}.$$

Fix a reduced word $w = x_{i_1}^{\varepsilon_1} \dots x_{i_k}^{\varepsilon_k} \in F(X_r)$. For $j \in [0, k]$ by w_j we denote the initial segment of w of length j . By e_j we denote the edge $w_{j-1} \xrightarrow{x_{i_j}^{\varepsilon_j}} w_j$ traversed by w in $\mathbf{Cay}(S_{r,d})$. By $\Gamma_{d,w}$ we denote the support graph for w in $\mathbf{Cay}(S_{r,d})$.

A word w is trivial in $S_{r,d}$ if and only if it defines the trivial flow π_w on $\mathbf{Cay}(S_{r,d-1})$. Obviously, the function π_w is trivial outside of the support graph $\Gamma_{d-1,w}$ and, hence:

$$w = 1 \text{ in } S_{r,d} \Leftrightarrow \pi_w \equiv 0 \text{ on } \Gamma_{d-1,w}.$$

Furthermore, for any $j \in [0, k]$ the graph Γ_{d-1,w_j} is a subgraph of $\Gamma_{d-1,w}$ and, hence, we can view π_{w_j} as a flow on $\Gamma_{d-1,w}$. In particular, $w_i = w_j$ in $S_{r,d}$ if and only if w_i and w_j define the same flows on $\Gamma_{d-1,w}$. The algorithm described in this section efficiently constructs graphs $\Gamma_{d,w}$ and flows π_w by induction on d . That goal is achieved via the concept of a distinguisher.

Definition 2.3. Let $w = w = x_{i_1}^{\varepsilon_1} \dots x_{i_k}^{\varepsilon_k} \in F(X_r)$. We say that a function $\nu_d : [0, k] \rightarrow [0, k]$ is a **distinguisher** for w in $S_{r,d}$ if it satisfies the following property:

$$\nu_d(i) = \nu_d(j) \Leftrightarrow w_i = w_j \text{ in } S_{r,d}.$$

A function $\epsilon_d : [1, k] \rightarrow [-k, k] \setminus \{0\}$ is called an **edge numbering function** for w in $S_{r,d}$ if:

- $\epsilon_d(i) = \epsilon_d(j)$ if and only if $e_i = e_j$;
- $\epsilon_d(i) = -\epsilon_d(j)$ if and only if $e_i = e_j^{-1}$. □

For any function $\nu : [0, k] \rightarrow [0, k]$ one can construct a rooted X -digraph $\Gamma_\nu = (V, E)$, with:

$$(2) \quad V = \{\nu(0), \dots, \nu(k)\} \text{ and } E = \{\nu(j-1) \xrightarrow{x_j} \nu(j) \mid j = 1, \dots, k\}$$

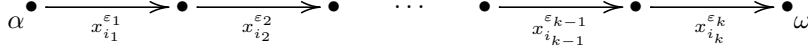
with the root at $\nu(0)$. It is easy to see that if ν_d is a distinguisher for w in $S_{r,d}$, then Γ_{ν_d} is isomorphic to the support graph $\Gamma_{d,w}$ and does not depend on a choice of a distinguisher ν_d .

Lemma 2.4. Given a distinguisher ν_d for w it requires quasi-linear time $\tilde{O}(|w|)$ to compute an edge-numbering function ϵ for w .

Proof. Each edge is uniquely defined by a triple $(\nu_d(j), \nu_d(j+1), x_{i_j}^{\varepsilon_j})$. As we explained in Section 1.6, we may assume that $r \leq |w|$. Hence, such triples can be encoded by bit-strings of length $O(\log_2 |w|)$. Organizing a tree of such bit-strings we can sort them and number lexicographically. Also, it is easy to check if two edges are inverses of each other. □

Our next goal is to construct a sequence of distinguishers ν_0, \dots, ν_d for a given w . Clearly, we can put $\nu_0 \equiv 0$ because $S_{r,0}$ is the trivial group. Assume that ν_{d-1} is constructed. Below we describe a procedure constructing a distinguisher ν_d .

Proposition 2.5. There exists a deterministic quasi-quadratic algorithm which for a word w and a distinguisher ν_{d-1} for w produces a distinguisher ν_d .

FIGURE 3. The graph $\Gamma(w)$.

Proof. Using Lemma 2.4 we number the edges of $\Gamma_{d-1,w}$ traversed by w in quasi-linear time. To construct ν_d we process w letter by letter constructing flows $\pi_{w_0}, \dots, \pi_{w_k}$. The function $\pi_{w_i} : E(\Gamma_{d-1,w}) \rightarrow \mathbb{Z}$ counts the algebraic number of times each edge is traversed by w_i . Since our edges are numbered we can encode π_{w_i} 's as tuples A_i of length $|E(\Gamma_{d-1,w})|$. The function π_{w_0} is encoded as the tuple of zeros. Clearly, A_{i+1} can be obtained from A_i by adding ± 1 to a single component. Each tuple A_i has length $|E(\Gamma_{d-1,w})| \leq |w|$ with absolute values of entries bounded by $|w|$. Hence, it takes $\tilde{O}(|w|)$ time to produce A_{i+1} from A_i . Thus, our procedure produces $|w| + 1$ bit-strings A_0, \dots, A_k of length $O(|w| \log_2 |w|)$ that uniquely represent the initial segments of w as elements of $S_{r,d}$. We organize these strings into a tree and number them according to the lexicographic order. The obtained numbering gives a required distinguisher ν_d .

It is straightforward to construct the tree described above. The size of the tree is $O(|w|^2 \log_2 |w|)$. Hence, the procedure has quasi-quadratic time complexity. \square

Theorem 2.6. *The word problem in $S_{r,d}$ can be solved by a deterministic procedure in quasi-quadratic time in $|w|$.*

Proof. Using the procedure described in the proof of Proposition 2.5 we compute distinguishers ν_0, \dots, ν_d for w . Computation of ν_{i+1} from ν_i requires quasi-quadratic time in $|w|$. It follows from Proposition 2.2 that if $d > \log_3 |w|$, then $w \neq 1$ in $S_{r,d}$. Hence, we only need to check the values of $d \leq \log_3 |w|$. Thus, we only need to compute up to $\log_3 |w|$ distinguishers. This implies that the procedure is quasi-quadratic in $|w|$. \square

This gives the first improvement to the algorithm described in [15].

3. ABSTRACT SUPPORT GRAPHS

In Section 2 we used support graphs to solve the word problem in free solvable groups in quasi-quadratic time. In Section 4 we design a randomized quasi-linear algorithm for the same problem. To better understand its behavior (to prove that it is false-biased) we need a notion of an abstract support graph. The basic idea is to forget that w is traced in $\mathbf{Cay}(S_{r,d})$ and consider any graph “covered” by w .

For a word $w = x_{i_1}^{\varepsilon_1} \dots x_{i_k}^{\varepsilon_k} \in F(X)$ define a rooted inverse X -digraph $\Gamma(w) = (V, E)$:

$$V = \{w_j \mid j \in [0, k]\} \quad \text{and} \quad E = \{w_{j-1} \xrightarrow{x_{i_j}^{\varepsilon_j}} w_j \mid j \in [1, k]\},$$

where $w_j = x_{i_1}^{\varepsilon_1} \dots x_{i_j}^{\varepsilon_j}$, see Figure 3. We say that a folded rooted X -digraph Γ is a support graph for w if there exists an X -digraph epimorphism $\pi : \Gamma(w) \rightarrow \Gamma$. Note that a morphism π is unique for Γ , because Γ is rooted and folded. Denote the set of all support graphs for w by Ω_w . Ω_w is the set all folded homomorphic images of $\Gamma(w)$. Hence, it is finite.

Let $\Gamma \in \Omega_w$. Since every initial segment of w defines a flow on Γ , we can define a graph $\iota(\Gamma)$:

$$V(\iota(\Gamma)) = \{\pi_{w_i} \mid i \in [0, k]\} \text{ and } E(\iota(\Gamma)) = \{\pi_{w_{j-1}} \xrightarrow{x_{i,j}^{\varepsilon_j}} \pi_{w_j} \mid j \in [1, k]\}.$$

It is easy to see that the map $w_i \xrightarrow{\pi} \pi_{w_i}$ defines an epimorphism $\pi : \Gamma(w) \rightarrow \iota(\Gamma)$, i.e., $\iota(\Gamma) \in \Omega_w$. Hence, the map $\Gamma \mapsto \iota(\Gamma)$ defines a function $\iota : \Omega_w \rightarrow \Omega_w$.

Proposition 3.1. *For any $\Gamma \in \Omega_w$ there exists a (unique) X -digraph epimorphism $\varphi : \iota(\Gamma) \rightarrow \Gamma$. Furthermore the following diagram commutes:*

$$\begin{array}{ccc} \Gamma(w) & \xrightarrow{\quad} & \iota(\Gamma) \\ & \searrow & \downarrow \varphi \\ & & \Gamma \end{array}$$

Proof. Vertices of $\iota(\Gamma)$ are flows π_{w_i} on Γ . Each flow π_{w_i} has the sink, which is the endpoint $\omega(p_{w_i})$ of the path p_{w_i} in Γ . Hence, we can define a map $\varphi : V(\iota(\Gamma)) \rightarrow V(\Gamma)$ by $\pi_{w_i} \mapsto \omega(p_{w_i})$. It is easy to check that φ is an X -digraph morphism satisfying

$$\pi_{w_i} \equiv \pi_{w_j} \Rightarrow \omega(p_{w_i}) = \omega(p_{w_j})$$

Therefore, the diagram indeed commutes. \square

Remark 3.2. *Let $H = \pi_1(\Gamma)$. The reader can recognize $\iota(\Gamma)$ as the image of $\Gamma(w)$ in the Schreier graph of the subgroup $[H, H] \leq F(X)$.* \square

The next proposition shows that a sequence of applications of ι always ends up with $\Gamma(w)$.

Lemma 3.3. *For any $\Gamma \in \Omega_w$ we have $\iota^{\log_3 |w|}(\Gamma) = \Gamma(w)$.*

Proof. If $\Gamma = \Gamma(w)$, then there is nothing to prove. Let m be the length of a shortest cycle in Γ . By Lemma 2.1, the length of a shortest cycle in $\iota(\Gamma)$ is not smaller than $3m$. Therefore, $\iota^{\log_3 |w|}(\Gamma)$ has no cycles, i.e., $\iota^{\log_3 |w|}(\Gamma) = \Gamma(w)$. \square

Lemma 3.4. *Let $\Gamma, \Delta \in \Omega_w$ and $\varphi : \Gamma \rightarrow \Delta$ be a rooted X -digraph morphism. Assume that u and v can be traced in Γ and define equal flows on Γ . Then they define equal flows on Δ . Therefore, if w defines the trivial flow on Γ , then it defines the trivial flow on Δ .*

Proof. Let $\pi_u^{(\Gamma)}, \pi_v^{(\Gamma)} : E(\Gamma) \rightarrow \mathbb{Z}$ and $\pi_u^{(\Delta)}, \pi_v^{(\Delta)} : E(\Delta) \rightarrow \mathbb{Z}$ be flows defined by u and v in Γ and in Δ respectively. Then for an arbitrary $e \in E(\Delta)$:

$$\pi_u^{(\Delta)}(e) = \sum_{f \in E(\Gamma), \varphi(f)=e} \pi_u^{(\Gamma)}(f) = \sum_{f \in E(\Gamma), \varphi(f)=e} \pi_v^{(\Gamma)}(f) = \pi_v^{(\Delta)}(e).$$

Hence, $\pi_u^{(\Delta)} \equiv \pi_v^{(\Delta)}$. By the same formula $\pi_u^{(\Gamma)} \equiv 0$ implies $\pi_u^{(\Delta)} \equiv 0$. \square

Proposition 3.5. *Let $\Gamma, \Delta \in \Omega_w$ and $\varphi : \Gamma \rightarrow \Delta$ be an epimorphism. Then there exists an epimorphism $\psi : \iota(\Gamma) \rightarrow \iota(\Delta)$ such that the diagram below commutes:*

$$\begin{array}{ccc} \iota(\Gamma) & \xrightarrow{\tau_1} & \Gamma \\ \downarrow \psi & & \downarrow \varphi \\ \iota(\Delta) & \xrightarrow{\tau_2} & \Delta \end{array}$$

Proof. By definition $V(\iota(\Gamma)) = \{\pi_{w_i}^{(\Gamma)} \mid i \in [0, k]\}$ and $V(\iota(\Delta)) = \{\pi_{w_i}^{(\Delta)} \mid i \in [0, k]\}$. The map $\psi : V(\iota(\Gamma)) \rightarrow V(\iota(\Delta))$ given by:

$$\pi_{w_i}^{(\Gamma)} \xrightarrow{\psi} \pi_{w_i}^{(\Delta)},$$

is well defined by Lemma 3.4. The map ψ takes an edge $\pi_{w_{i-1}}^{(\Gamma)} \xrightarrow{x_i} \pi_{w_i}^{(\Gamma)}$ in Γ to the edge $\pi_{w_{i-1}}^{(\Delta)} \xrightarrow{x_i} \pi_{w_i}^{(\Delta)}$ in Δ . Therefore, ψ preserves connectedness and labels and is indeed an X -digraph epimorphism.

Finally we note that for any $\pi_{w_{i-1}}^{(\Gamma)} \in V(\iota(\Gamma))$ we have $\tau_1(\pi_{w_{i-1}}^{(\Gamma)})$ is the endpoint of w_i traced in Γ . Similarly, $\tau_2(\psi(\tau_1(\pi_{w_{i-1}}^{(\Gamma)})))$ is the endpoint of w_i traced in Δ . Since, φ is an X -digraph morphism taking the root to the root, we have a commuting diagram. \square

3.1. Language support graphs. Definition of a word support graph can be generalized to any set of words $S \subseteq F(X)$ as follows. Define a *prefix tree* $T = T(S)$:

$$V(T) = \{u \in F(X) \mid u \circ v \in S\} \text{ and } E(T) = \{w \xrightarrow{x} wx \mid w, wx \in S, x \in X^\pm\}.$$

We say that a rooted inverse X -digraph Γ is a *support graph* for S if there exists an X -digraph epimorphism $T \rightarrow \Gamma$. Assume that S is finite. The (finite) set of all support graphs for S is denoted by Ω_S . For any $\Gamma \in \Omega_S$ we can define the graph $\iota(\Gamma) = (V, E)$:

$$V(\iota(\Gamma)) = \{\pi_w \mid w \in V(T(S))\} \text{ and } E(\iota(\Gamma)) = \{\pi_w \xrightarrow{x} \pi_{wx} \mid w, wx \in V(T(S))\}.$$

It is easy to check that all results in this section for word support graphs hold for language support graph as well. Lemma 3.3 can be reformulated as follows:

Lemma 3.6. *Let S be a finite subset of $F(X_r)$ and d the diameter of the tree $T(S)$. Then for any $\Gamma \in \Omega_S$ we have $\iota^{\log_3 d}(\Gamma) = T(S)$.* \square

4. THE WORD PROBLEM: RANDOMIZED SOLUTION

In this section we improve quasi-quadratic procedure described in Proposition 2.5, we make it quasi-linear. Let $w = x_{i_1}^{\varepsilon_1} \dots x_{i_k}^{\varepsilon_k} \in F(X)$, Γ_{d-1} be the support graph for w in $S_{r,d-1}$, and $m = |E(\Gamma_{d-1})|$. Recall that the algorithm computes the set of tuples $A_0, \dots, A_k \in \mathbb{Z}^m$ that represent the flows $\pi_{w_0}, \dots, \pi_{w_k}$ on Γ_{d-1} . Tuples A_i are further encoded as bit-strings of lengths $O(|w| \log_2 |w|)$. Hence, we deal with $|w|$ objects of size $|w|$ which makes quadratic complexity.

To improve complexity we choose a point $A(a_1, \dots, a_m) \in [0, |w|^3]^m$ uniformly randomly and replace A_i 's with the numbers $d^2(A, A_i)$, where d is the Euclidean distance in \mathbb{Z}^m . Those numbers have bit-lengths $O(\log_2 |w|)$. The next two lemmas are concerned with complexity of computing the numbers $d^2(A, A_i)$.

Lemma 4.1. *It requires $\tilde{O}(|w|)$ time to compute $d^2(A, A_0) = \sum_{i=1}^m a_i^2$.*

Proof. For every $i = 1, \dots, |w|$ the bit-length of a_i is bounded by $3 \log |w|$. Schönhage-Strassen algorithm requires $O(\log |w| \log \log |w| \log \log \log |w|)$ steps to compute each a_i^2 . The bit-length of a_i^2 is bounded by $6 \log |w|$. Finally, it requires $O(|w| \log |w|)$ steps to sum m obtained squares each of length $6 \log |w|$. Thus, the total complexity is $O(|w| \log |w| \log \log |w| \log \log \log |w|)$. \square

Let $\overline{AA_i} = (A_{i,1}, \dots, A_{i,m})$. The vectors $\overline{AA_i}$ and $\overline{AA_{i+1}}$ differ at a single, say j th, component and $|A_{i,j} - A_{i+1,j}| = 1$. Therefore,

$$(3) \quad d^2(A, A_{i+1}) - d^2(A, A_i) = \begin{cases} 2|A_{i,j}| + 1 & \text{if } |A_{i+1,j}| > |A_{i,j}|, \\ -2|A_{i,j}| + 1 & \text{if } |A_{i+1,j}| < |A_{i,j}|. \end{cases}$$

Lemma 4.2. *Given $\overline{AA_i}$ and $d^2(A, A_i)$ it takes $O(\log |w|)$ time to compute $\overline{AA_{i+1}}$ and $d^2(A, A_{i+1})$.*

Proof. First note that

$$d^2(A, A_i) \leq \sum_{i=1}^m (|w|^3 + |w|)^2 \leq (|w|^3 + |w|)^2 |w|.$$

Hence, $d^2(A, A_i)$ has bit-length $O(\log |w|)$. Similarly, $\pm 2|A_{i,j}| + 1$ has bit-length $O(\log |w|)$. It requires $O(\log |w|)$ to compute $\pm 2|A_{i,j}| + 1$. Finally, it takes the same time to take the sum of $d^2(A, A_i)$ and $\pm 2|A_{i,j}| + 1$. \square

Proposition 4.3. *There exists a randomized quasi-linear algorithm which given a word $w = x_{i_1}^{\varepsilon_1} \dots x_{i_k}^{\varepsilon_k} \in F(X_r)$ and a distinguisher ν_{d-1} for w produces a function ν_d . The function ν_d is a distinguisher for w with the probability at least $1 - \frac{1}{|w|}$, where the probability is taken over all uniform choices of the point A in $[0, |w|^3]^m$.*

Proof. Using Lemmas 4.1 and 4.2 we can compute the array $i \mapsto d^2(A, A_i)$ in $\tilde{O}(|w|)$ time. The numbers $\{d(A, A_i)\}_{i=0}^{|w|}$ are then lexicographically sorted and numbered $d^2(A, A_i) \mapsto n_i$. The function ν_d is the composition $i \mapsto n_i$. Overall, it requires in $\tilde{O}(|w|)$ time to compute ν_d .

The described algorithm makes a mistake when $d^2(A, A_i) = d^2(A, A_j)$ while $A_i \neq A_j$ for some $i, j \in [0, |w|]$. This happens only when the randomly chosen point A belongs to a hyperplane with a normal vector $\overline{A_i A_j}$ through $\frac{1}{2}(A_i + A_j)$. The union of hyperplanes for all pairs of points (A_i, A_j) contains at most $\frac{1}{|w|}$ part of the hypercube $[0, |w|^3]^m$. Hence, the uniformly chosen A has not greater than $\frac{1}{|w|}$ chance to collapse two distinct points A_i, A_j . \square

The next proposition states that the support graph of the function produced by our algorithm is a homomorphic image of the correct support graph $\Gamma_{d,w}$.

Proposition 4.4. *Let ν_{d-1}, ν_d be distinguishers for w , ν'_d be a function produced from ν_{d-1} by the randomized procedure described in Proposition 4.3, and $\Delta = \Gamma_{\nu'_d}$ (see formulae (2)). Then there exists an epimorphism $\varphi : \Gamma_d \rightarrow \Delta$ which is an isomorphism if and only if the algorithm does not make a mistake.*

Proof. For any $i, j \in [0, k]$:

$$\nu_d(i) = \nu_d(j) \Leftrightarrow A_i = A_j \Rightarrow \nu'_d(i) = \nu'_d(j).$$

Therefore, there exists a (unique) epimorphism $\varphi : \Gamma_d \rightarrow \Delta$. Clearly, φ is an isomorphism if and only if

$$\nu_d(i) = \nu_d(j) \Leftrightarrow \nu'_d(i) = \nu'_d(j),$$

i.e., when the algorithm outputs a correct distinguisher. \square

Theorem 4.5. *Let $r, d \in \mathbb{N}$ and $w \in F(X_r)$. There exists a quasi-linear randomized algorithm deciding if $w = 1$ in $S_{r,d}$, or not. Furthermore,*

- (a) *if $w = 1$ in $S_{r,d}$, then the algorithm outputs Yes;*
- (b) *if $w \neq 1$ in $S_{r,d}$, then the algorithm outputs No with probability at least*

$$\left(1 - \frac{1}{|w|}\right)^{\log_3 |w|}.$$

Proof. Starting with $\nu_0 \equiv 0$ we compute distinguishers ν'_0, \dots, ν'_d using the randomized algorithm described in the proof of Proposition 4.3. Output *Yes* if $\nu'_d(|w|) = \nu'_d(0)$. Otherwise, output *No*. Since d can be bounded by $\log_3 |w|$ the described algorithm has quasi-linear complexity in $|w|$.

Assume that $w = 1$ in $S_{r,d}$. Let ν'_1, \dots, ν'_d be the sequence of functions inductively produced by the randomized algorithm. By construction we have $\nu'_0 = \nu_0 \equiv 0$. Denote $\Gamma_{\nu'_i}$ by Δ_i . It follows from Propositions 4.4 and 3.5 that for every $i \in [0, k]$ there exists an epimorphism $\tau_i : \Gamma_i \rightarrow \Delta_i$. Since w is trivial in $S_{r,d}$, then it has the trivial flow in Γ_{d-1} . Hence, by Lemma 3.4, w also has the trivial flow in Δ_{d-1} . Therefore, the algorithm outputs *Yes*.

We compute up to $\log_3 |w|$ distinguishers. By Proposition 4.3, the chance to make a mistake at any stage is not greater than $\frac{1}{|w|}$. Thus, the chance to make no mistakes is not smaller than $\left(1 - \frac{1}{|w|}\right)^{\log_3 |w|}$. \square

Finally, we want to make several remarks on the performance of the algorithm. The bound in Theorem 4.5(b) can be simplified as follows:

$$P(\text{success}) \geq \left(1 - \frac{1}{|w|}\right)^{\log_3 |w|} \geq 1 - \frac{\log_3 |w|}{|w|}.$$

Thus, the failure rate of the algorithm decreases almost linearly with $|w|$.

The success rate of the algorithm can be improved by sampling the point A from a bigger hypercube. For instance, taking numbers uniformly from $[0, |w|^4]$ improves the correctness estimate in Proposition 4.3 to $1 - \frac{1}{|w|^2}$ and the correctness estimate in Theorem 4.5 to $1 - \frac{\log_3 |w|}{|w|^2}$. At the same time bit-lengths of $d^2(A, A_i)$ increase only by a constant factor leaving the quasi-linear complexity bound intact.

The actual correctness probability is probably much better than our estimates. Making a mistake on some intermediate step does not imply that the algorithm will output *Yes* on $w \neq 1$. In fact, it is possible to get correct distinguisher ν_{i+1} starting from incorrect ν_i .

5. THE POWER PROBLEM

In this section we describe the algorithm for solving the power problem in $S_{r,d}$. The algorithm is based on two observations. The first observation is:

$$u = v^k \text{ in } S_{r,d} \Rightarrow u = v^k \text{ in } S_{r,d-1}.$$

The second observation is the Malcev theorem on centralizers in free solvable groups.

Theorem ([13]). *The centralizer of any nontrivial element u of a free solvable group is abelian. Furthermore:*

- (a) *If $u = 1$ in $S_{r,d-1}$, then $[u, v] = 1$ in $S_{r,d}$ if and only if $v = 1$ in $S_{r,d-1}$.*
- (b) *If $u \neq 1$ in $S_{r,d-1}$, then $[u, v] = 1$ in $S_{r,d}$ if and only if u and v are powers of the same element in $S_{r,d}$.* \square

The next algorithm solves the power problem in $S_{r,d}$.

Power problem in $S_{r,d}$

Require: $r, d \in \mathbb{N}$ and $u, v \in F(X_r)$.

Ensure: $k \in \mathbb{Z}$ such that $u = v^k$ in $S_{r,d}$ and *Fail* if k does not exist.

- 1: Let $D = 1 + \min\{d, \log_3(|u| + |v|)\}$.
 - 2: Construct a sequence of support graphs $\{\Gamma_i\}_{i=0}^D$ for the set $\{u, v, [u, v]\}$.
 - 3: Use $\{\Gamma_i\}_{i=0}^D$ to compute the largest s such that $u = 1$ in $S_{r,s}$.
 - 4: Use $\{\Gamma_i\}_{i=0}^D$ to compute the largest t such that $v = 1$ in $S_{r,t}$.
 - 5: **if** $d \leq s, t$ **then return** 1.
 - 6: **if** $s < d \leq t$ **then return** *Fail*.
 - 7: **if** $t < d \leq s$ **then return** 0.
 - 8: **if** $s < t < d$ **then return** *Fail*.
 - 9: **if** $t < s < d$ **then return** *Fail*.
 - 10: **if** $s = t < d$ **then**
 - 11: Choose any edge e in Γ_s with $\pi_v(e) \neq 0$ in Γ_s and put $k = \pi_u(e)/\pi_v(e)$.
 - 12: **if** $k \notin \mathbb{Z}$ or $[u, v] \neq 1$ in $S_{r,d}$ **then return** *Fail*.
 - 13: **for all** e in Γ_s **do**
 - 14: **if** $\pi_u(e) \neq k\pi_v(e)$ **then return** *Fail*.
 - 15: **end for**
 - 16: **return** k .
 - 17: **end if**
-

A few details are in order. By Lemma 3.6 the support graph for $T = T(\{u, v, [u, v]\})$ in $S_{r,1+\log_3(|u|+|v|)}$ is T itself because the diameter of the graph T is not greater than $3(|u| + |v|)$. In particular, $s, t \leq 1 + \log_3(|u| + |v|)$. That explains the choice of D .

Algorithm 6.2 can be implemented as a deterministic or a randomized algorithm depending on how we compute the sequence of graphs $\{\Gamma_i\}_{i=0}^D$. Computing $\{\Gamma_i\}_{i=0}^D$ using the deterministic algorithm from Theorem 2.6 gives the deterministic version of Algorithm 5.

Theorem 5.1. *The deterministic Algorithm 5 solves the power problem in $S_{r,d}$ in quasi-quadratic time $\tilde{O}((|u| + |v|)^2)$.*

Proof. All cases considered in the algorithm are trivial except, maybe, the case when $s = t < d$. In that case $u \neq 1$ and $v \neq 1$ in $S_{r,d}$. The flows π_u and π_v are circulations in $\mathbf{Cay}(S_{r,s})$ (the source and the sink are the same) and $v = u^k$ implies that $\pi_v \equiv k\pi_u$. Hence, $\pi_u(e)/\pi_v(e)$ is the only possible value for k (if $\pi_v(e) \neq 0$). Now, we have two cases as in the Malcev theorem. If $s = t = d - 1$, then it is sufficient to check if $\pi_v \equiv k\pi_u$ (done in lines 13–15). Otherwise, it is sufficient to check if $[u, v] = 1$ (part (b) of the Malcev theorem) which is done in line 12. By Theorem 2.6 it takes quasi-quadratic time to construct support graphs for $T(\{u, v, [u, v]\})$ and test the equality $[u, v] = 1$. \square

Computing the sequence $\{\Gamma_i\}_{i=0}^D$ can also be done using the randomized algorithm from Theorem 4.5. To obtain the desired probability of success we choose the random tuple A with components chosen uniformly from $[0, 9(|u| + |v|)^3]$.

Theorem 5.2. *The randomized Algorithm 5 solves the power problem in $S_{r,d}$ in quasi-linear time $\tilde{O}(|u| + |v|)$. The algorithm returns a correct answer with probability at least:*

$$\left(1 - \frac{1}{|u|+|v|}\right)^{1+\log_3(|u|+|v|)}.$$

Proof. The complexity estimate immediately follows from Theorem 4.5. We argue as in Proposition 4.3 to get the correctness lower-bound. The graph $\Gamma(\{u, v, [u, v]\})$ has at most $3(|u| + |v|)$ vertices which defines at most $9(|u| + |v|)^2$ bad hyperplanes. The union of those hyperplanes can contain at most $\frac{1}{|u|+|v|}$ part of our hypercube. Hence, our procedure produces the correct support graph Γ_i for $\Gamma(\{u, v, [u, v]\})$ in $S_{r,i}$ with probability at least $1 - \frac{1}{|u|+|v|}$. We perform up to $1 + \log_3(|u| + |v|)$ iterations. Hence the claimed correctness probability. \square

Algorithm 5 is unbiased, i.e., it can make an error on both positive and negative instances of the problem.

6. THE CONJUGACY PROBLEM

In this section we revisit algorithmic difficulty of the conjugacy problem in free solvable groups. In [14] Matthews proved that the conjugacy problem (CP) is solvable in wreath products $A \mathbf{wr} B$ (under some natural assumptions on A and B). She used that result to prove that CP in free metabelian groups is decidable. Kargapolov and Remeslennikov generalized that result to free solvable groups in [9]. A few years later Remeslennikov and Sokolov in [20] described precisely the image of $F/[N, N]$ under the Magnus embedding and showed that two elements are conjugate in $S_{r,d}$ if and only if their images are conjugate in $M(F/N)$. Recently Vassileva in [24] found a polynomial time $O(rd(|u| + |v|)^8)$ algorithm for the conjugacy problem combining the Matthews and Remeslennikov-Sokolov results.

6.1. Matthews algorithm for wreath products. In this section we shortly outline computations in the proof of the Matthews theorem on conjugacy in wreath products. Note that we use different notation for wreath products than Matthews and at the end we obtain slightly different formula.

Let A, B be finitely generated groups. By A^B we denote the set of all functions $f : B \rightarrow A$ with finite support. For $f \in A^B$ and $\alpha \in B$ define $f^\alpha \in A^B$ as follows:

$$f^\alpha(x) = f(\alpha^{-1}x).$$

The *restricted wreath product* of A and B , denoted by $A \mathbf{wr} B$, is a set of pairs:

$$\{(f, \alpha) \mid f \in A^B, \alpha \in B\}$$

with multiplication given by:

$$(f, \alpha) \cdot (g, \beta) = (fg^\alpha, \alpha\beta).$$

Hence, for $x = (f, \alpha)$, $y = (g, \beta)$, and $z = (h, \gamma)$ in $A \mathbf{wr} B$ we have:

$$zx = yz \Leftrightarrow \begin{cases} \gamma\alpha = \beta\gamma & \text{in } B, \\ f^\gamma = h^{-1}gh^\beta & \text{in } A^B. \end{cases}$$

Assuming the equality above we observe that for any $\delta \in B$ and $j \in \mathbb{Z}$:

$$\begin{aligned} f(\gamma^{-1}\beta^j\delta) &= f^\gamma(\beta^j\delta) \\ &= h^{-1}(\beta^j\delta)g(\beta^j\delta)h^\beta(\beta^j\delta) \\ &= h^{-1}(\beta^j\delta)g(\beta^j\delta)h(\beta^{j-1}\delta) \end{aligned}$$

and, hence, the following formula holds for any $m \in \mathbb{Z}$ and $n \in \mathbb{N}$:

$$(4) \quad \prod_{j=m}^{m-n} f^\gamma(\beta^j\delta) = h^{-1}(\beta^m\delta) \cdot \left(\prod_{j=m}^{m-n} g(\beta^j\delta) \right) \cdot h(\beta^{m-n-1}\delta).$$

Now, for $\beta, \gamma, \delta \in B$ and $f : B \rightarrow A$ define $\rho^{(f, \beta, \gamma)}(\delta) \in A$ as follows:

$$\rho^{(f, \beta, \gamma)}(\delta) = \begin{cases} \prod_{j=k-1}^0 f(\gamma^{-1}\beta^j\delta) & \text{if } \mathbf{ord}(\beta) = k, \\ \prod_{j=-\infty}^{-1} f(\gamma^{-1}\beta^j\delta) & \text{if } \mathbf{ord}(\beta) = \infty. \end{cases}$$

Hence, assuming that $zx = yx$ and using the defined above notation, equality (4) gives:

- if $\mathbf{ord}(\beta) < \infty$, then $\rho^{(f, \beta, \gamma)}(\delta)$ is conjugate to $\rho^{(g, \beta, 1)}(\delta)$ in A for every $\delta \in B$;
- if $\mathbf{ord}(\beta) = \infty$, then $\rho^{(f, \beta, \gamma)}(\delta) = \rho^{(g, \beta, 1)}(\delta)$ in A for every $\delta \in B$.

Matthews proved in [14] that the converse also holds.

Theorem 6.1 ([14]). *Let A and B be finitely generated groups. Elements $x = (f, \alpha)$ and $y = (g, \beta)$ are conjugate in $A \mathbf{wr} B$ if and only if there exists $\gamma \in B$ satisfying:*

- (a) $\gamma\alpha = \beta\gamma$ in B ;
- (b) if $\mathbf{ord}(\beta) < \infty$, then $\rho^{(f, \beta, \gamma)}(\delta)$ is conjugate to $\rho^{(g, \beta, 1)}(\delta)$ in A for every $\delta \in B$;
- (c) if $\mathbf{ord}(\beta) = \infty$, then $\rho^{(f, \beta, \gamma)}(\delta) = \rho^{(g, \beta, 1)}(\delta)$ in A for every $\delta \in B$. \square

Recall that the Magnus embedding embeds $S_{r,d}$ into $\mathbb{Z}^r \mathbf{wr} S_{r,d-1}$. The Matthews theorem gives a solution to the conjugacy problem in $\mathbb{Z}^r \mathbf{wr} S_{r,d-1}$ and Remeslennikov-Sokolov prove that elements are conjugate in $S_{r,d}$ if and only if their images are conjugate in $\mathbb{Z}^r \mathbf{wr} S_{r,d-1}$. This solves the conjugacy in $S_{r,d}$ and concludes the general algorithm description.

6.2. Geometric approach to conjugacy problem in free solvable groups.

In the case of a free solvable group the Matthews result can be formulated in a geometric way using flows on Schreier graphs. By $\mathbf{Sch}_d(y)$ we denote the Schreier graph of $\langle y \rangle$ in $S_{r,d}$. The next lemma follows from the definition of ρ .

Lemma 6.2. *Let $\phi : S_{r,d} \rightarrow \mathbb{Z}^r \mathbf{wr} S_{r,d-1}$ be the Magnus embedding. Let $x, y, z \in F(X_r)$ and $\phi(x) = (f, \alpha)$, $\phi(y) = (g, \beta)$, $\phi(z) = (h, \gamma)$. Then for any $\delta \in S_{r,d-1}$:*

- (a) $\rho^{(g, \beta, 1)}(\delta) \in \mathbb{Z}^r$ is exactly the value of π_y in $\mathbf{Sch}_{d-1}(y)$ restricted to the edges $\delta \rightarrow \delta x_i$ for $i = 1, \dots, r$;
- (b) $\rho^{(f, \beta, \gamma)}(\delta) \in \mathbb{Z}^r$ is exactly the value of $\pi_{\gamma x \gamma^{-1}}$ in $\mathbf{Sch}_{d-1}(y)$ restricted to the edges $\delta \rightarrow \delta x_i$ for $i = 1, \dots, r$. \square

Lemma 6.3. *For any $y \in F(X_r)$ we have $\pi_y \equiv 0$ in $\mathbf{Sch}_{d-1}(y)$ if and only if $y = 1$ in $S_{r,d}$.*

Proof. The equality $\pi_y \equiv 0$, by Lemma 6.2, implies $\rho \equiv 0$. Hence, by Theorem 6.1, $y \sim 1$ in $S_{r,d}$, i.e., $y = 1$ in $S_{r,d}$.

Conversely, if $y = 1$ in $S_{r,d}$, then $y = 1$ in $S_{r,d-1}$. Hence, $\mathbf{Sch}_{d-1}(y) = \mathbf{Cay}(S_{r,d-1})$ and $\pi_y \equiv 0$ in $\mathbf{Cay}(S_{r,d-1})$. Hence, sufficiency holds. \square

The next theorem states that $x \sim y$ in $S_{r,d}$ if and only if there exists a shift of x in $\mathbf{Sch}_{d-1}(y)$ defining the same flow as y .

Theorem 6.4. *Words $x, y \in F(X_r)$ represent conjugate elements in $S_{r,d}$ if and only if there exists $z \in F(X)$ satisfying $\pi_{zxz^{-1}} \equiv \pi_y$ in $\mathbf{Sch}_{d-1}(y)$. The element z can be viewed as an element of $S_{r,d-1}$.*

Proof. If x and y represent conjugate elements in $S_{r,d}$, then for some z we have $zxz^{-1} = y$ in $S_{r,d}$. Hence, $\pi_{zxz^{-1}} \equiv \pi_y$ in $\mathbf{Cay}(S_{r,d-1})$ and in $\mathbf{Sch}_{d-1}(y)$. Thus, necessity holds.

Conversely, assume that there exists $z \in F(X)$ satisfying $\pi_{zxz^{-1}} \equiv \pi_y$ in $\mathbf{Sch}_{d-1}(y)$. If $\pi_y \equiv 0$ in $\mathbf{Sch}_{d-1}(y)$, then, by Lemma 6.3, $y = 1$ in $S_{r,d}$ and $\mathbf{Sch}_{d-1}(y) = \mathbf{Cay}(S_{r,d-1})$. Hence, $\pi_{zxz^{-1}} \equiv 0$ in $\mathbf{Cay}(S_{r,d-1})$ and $z = 1$ in $S_{r,d}$ as well.

Hence, we may assume that $\pi_y \not\equiv 0$ in $\mathbf{Sch}_{d-1}(y)$. Now the equality $\pi_{zxz^{-1}} \equiv \pi_y$ in $\mathbf{Sch}_{d-1}(y)$ implies that zxz^{-1} is a cycle in $\mathbf{Sch}_{d-1}(y)$, i.e., zxz^{-1} is a power of y . Furthermore, since $\pi_y \not\equiv 0$, we clearly have $zxz^{-1} = y$ in $S_{r,d-1}$. Thus, item (a) of Theorem 6.1 holds.

Let $\phi(x) = (f, \alpha)$, $\phi(y) = (g, \beta)$, $\phi(z) = (h, \gamma)$. By Lemma 6.2 the equality $\pi_{zxz^{-1}} \equiv \pi_y$ implies that $\rho^{(f, \beta, \gamma)}(\delta) = \rho^{(g, \beta, 1)}(\delta)$ for every $\delta \in S_{r,d-1}$. Hence, both items (b) and (c) of Theorem 6.1 hold. \square

Now it is straightforward to solve the conjugacy problem in $S_{r,d}$.

Conjugacy problem in $S_{r,d}$

Require: $r, d \in \mathbb{N}$ and $x, y \in F(X_r)$.

Ensure: *Yes* if $u \sim v$, *No* otherwise.

- 1: **if** $x = 1$ and $y = 1$ **then return** *Yes*.
 - 2: **if** $x = 1$ or $y = 1$ **then return** *No*.
 - 3: Using Algorithm 5 construct the support graph for y in $\mathbf{Sch}_{d-1}(y)$ and the flow π_y .
 - 4: Choose a prefix y_i satisfying $\pi_y(y_i \rightarrow y_{i+1}) \neq 0$.
 - 5: **for all** x' such that $x = x' \circ x''$ **do**
 - 6: Put $\gamma = y_i x'^{-1}$.
 - 7: Using Algorithm 5 compute the flow $\pi_{\gamma x \gamma^{-1}}$ in $\mathbf{Sch}_{d-1}(y)$.
 - 8: **if** $\pi_{\gamma x \gamma^{-1}} \equiv \pi_y$ **then return** *Yes*.
 - 9: **end for**
 - 10: **return** *No*.
-

A few details are in order. To construct support graphs for y and $\gamma x \gamma^{-1}$ in $\mathbf{Sch}_{d-1}(y)$ one needs to find all prefixes of y and $\gamma x \gamma^{-1}$ define the same $\langle y \rangle$ -cosets. The later problem reduces to the membership problem for $\langle y \rangle$ and can be treated by Algorithm 5. It follows from the choice of γ 's that the inputs to Algorithm 5 have lengths bounded by $|x| + |y|$.

Theorem 6.5. *There exists a quasi-quintic time $\tilde{O}((|x| + |y|)^5)$ deterministic algorithm solving the conjugacy problem in $S_{r,d}$.*

Proof. Correctness of the algorithm follows from Theorem 6.4. The loop 5–9 performs $|x|+1$ iterations. At each iteration we compute the flow $\pi_{\gamma x \gamma^{-1}}$ which requires $O(|x|^2)$ runs of Algorithm 5 and test if $\pi_{\gamma x \gamma^{-1}} \equiv \pi_y$ which requires $O((|x|+|y|)^2)$ more runs of Algorithm 5. The deterministic Algorithm 5 has quasi-quadratic time complexity. Thus, the total time complexity is $\tilde{O}((|x|+|y|)^5)$. \square

We can further improve efficiency if we use the randomized version of Algorithm 5. Algorithm 6.2 invokes Algorithm 5 at most $(|x|+|y|)^2$ times on each iteration, hence the total number of invocations is bounded by $(|x|+|y|)^3$. Each invocation of Algorithm 5 can produce an incorrect answer. To better control the error we go deep into details of Algorithm 5 again. As we mentioned above the lengths of inputs (u, v) for Algorithm 5 are bounded by $|x|+|y|$. Hence, for $T = T(\{u, v, [u, v]\})$ we have:

$$\text{diam}(T) \leq |u| + |v| + |[u, v]| \leq 5(|x| + |y|).$$

Therefore, every time randomized Algorithm 5 is invoked it performs at most $\log_3(5(|x|+|y|)) \leq 2 + \log_3(|x|+|y|)$ iterations. The number of vertices in T is also bounded by $5(|x|+|y|)$. Hence, the total number of bad hyperplanes is not greater than $25(|x|+|y|)^2$. Therefore, choosing a random tuple A with elements in $[0, 25(|x|+|y|)^6]$ produces the correct result on a single iteration with probability not less than

$$1 - \frac{25(|x|+|y|)^2}{25(|x|+|y|)^6} \leq 1 - \frac{1}{(|x|+|y|)^4}.$$

Hence, we get the correct result on a single invocation of an algorithm 5 with probability at least:

$$\left(1 - \frac{1}{(|x|+|y|)^4}\right)^{2+\log_3(|x|+|y|)}.$$

Performing $(|x|+|y|)^3$ invocations of Algorithm 5 results in all correct results with probability at least

$$\left(\left(1 - \frac{1}{(|x|+|y|)^4}\right)^{2+\log_3(|x|+|y|)}\right)^{(|x|+|y|)^3} \geq \left(1 - \frac{1}{|x|+|y|}\right)^{2+\log_3(|x|+|y|)}.$$

This proves the following theorem.

Theorem 6.6. *There exists a quasi-quadratic time $\tilde{O}((|x|+|y|)^4)$ unbiased randomized algorithm solving the conjugacy problem in $S_{r,d}$. The probability of a correct computation is at least $\left(1 - \frac{1}{|x|+|y|}\right)^{2+\log_3(|x|+|y|)}$.* \square

REFERENCES

- [1] M. Agrawal, N. Kayal, and N. Saxena, *PRIMES is in P*, Ann. of Math. 160 (2004), pp. 781–793.
- [2] L. Babai, *Monte-Carlo algorithms in graph isomorphism testing*, Université de Montréal, D.M.S. No. 79–10, 1979.
- [3] K. T. Chen, R. H. Fox, and R. C. Lyndon, *Free differential calculus IV*, Ann. of Math. 71 (1960), pp. 408–422.
- [4] C. Droms, J. Lewin, and H. Servatius, *The length of elements in free solvable groups*, Proc. Amer. Math. Soc. 119 (1993), pp. 27–33.
- [5] R. H. Fox, *Free differential calculus I*, Ann. of Math. 57 (1953), pp. 547–560.
- [6] ———, *Free differential calculus II*, Ann. of Math. 59 (1954), pp. 196–210.
- [7] ———, *Free differential calculus III*, Ann. of Math. 64 (1956), pp. 407–419.
- [8] I. Kapovich and A. G. Miasnikov, *Stallings foldings and subgroups of free groups*, J. Algebra 248 (2002), pp. 608–668.

- [9] M. I. Kargapolov and V. N. Remeslennikov, *The conjugacy problem for free solvable groups*, Algebra i Logika Sem. 5 (1966), pp. 15–25. (Russian).
- [10] R. Lyndon and P. Schupp, *Combinatorial Group Theory*, Classics in Mathematics. Springer, 2001.
- [11] W. Magnus, *On a theorem of Marshall Hall*, Ann. of Math. 40 (1939), pp. 764–768.
- [12] W. Magnus, A. Karrass, and D. Solitar, *Combinatorial Group Theory*. Springer-Verlag, 1977.
- [13] A. Malcev, *On free solvable groups*, Soviet Math. Doklady 1 (1960), pp. 65–68.
- [14] J. Matthews, *The conjugacy problem in wreath products and free metabelian groups*, Trans. Amer. Math. Soc. 121 (1966), pp. 329–339.
- [15] A. G. Miasnikov, V. Romankov, A. Ushakov, and A. Vershik, *The word and geodesic problems in free solvable groups*, Trans. Amer. Math. Soc. 362 (2010), pp. 4655–4682.
- [16] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge University Press, 1995.
- [17] A. Naik, K. Regan, and D. Sivakumar, *On quasilinear-time complexity theory*, Theoret. Comput. Sci. 148 (1995), pp. 325–349.
- [18] M. Rabin, *Probabilistic algorithms*. Algorithms and complexity: new directions and recent results, pp. 21–39. Academic Press, 1976.
- [19] ———, *Probabilistic tests for primality*, J. of Number Theory 12 (1980), pp. 128–138.
- [20] V. N. Remeslennikov and V. G. Sokolov, *Certain properties of Magnus embedding*, Algebra i Logika 9 (1970), pp. 566–578.
- [21] A. Schönhage and V. Strassen, *Schnelle multiplikation großer zahlen*, Computing 7 (1971), pp. 281–292.
- [22] R. Solovay and V. Strassen, *A fast Monte-Carlo test for primality*, SIAM J. Comput. 6 (1977), pp. 84–85.
- [23] J. Stallings, *Topology of finite graphs*, Invent. Math. 71 (1983), pp. 551–565.
- [24] S. Vassilieva, *Polynomial time conjugacy in wreath products and free solvable groups*, Groups Complex. Cryptol. 3 (2011), pp. 105–120.
- [25] A. M. Vershik and S. Dobrynin, *Geometrical approach to the free solvable groups*, Int. J. Algebra Comput. 15 (2005), pp. 1243–1260.

STEVENS INSTITUTE OF TECHNOLOGY, HOBOKEN, NJ, 07030 USA

E-mail address: aushakov@stevens.edu